IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 2, Ser. III (Mar - Apr 2019), PP 06-11
www.iosrjournals.org

Overcoming the Drawbacks of Convolutional Neural Network Using Capsule Network

Shraddha More[1], Vrushali Shirodkar [2], Raunak Joshi[3], Nrupali Thakur[4]
[1]*(St. John College of Engineering & Management, India)*
[2]*(St. John College of Engineering & Management, India)*
[3]*(St. John College of Engineering & Management, India)*
[4]*(St. John College of Engineering & Management, India)*
Corresponding Author Shraddha More:

Abstract : *Convolutional Neural Networks were inspired by biological processes that have proven very effective in areas such as image recognition and classification. CNNs have been successful in identifying faces, objects, things and traffic signs. But at a certain point there are some disadvantages of neural networks which we are going to resolve using Capsules. Capsules at one level make predictions, via transformation matrices and using other algorithms. When multiple of predictions agree, a higher level capsule becomes active. A Capsule is a group of neurons whose vector represents the parameter instance of a very specific type of entity which can be termed as an object or a part of an object. The length of an activity vector is used to represent the probability that an entity exists and its orientation to represent the instantiation parameters. The problems in existing method are that the inverted images or images with noise cannot be detected. But by using our proposed system the drawbacks of the present system will be overcome. The aim of our proposed system is to overcome the disadvantages of convolutional Neural Networks Using Capsules.*

Keywords: *Image- Processing, Deep Learning, GPU, Image Classification*

Date of Submission: 27-03-2019 Date of acceptance: 11-04-2019

I. Introduction

Over the course of years there is been development in the field of Convolutional Neural Networks. However the years of advancements failed in the some areas. These areas include spatial coordinate errors, training and detection in different sample spaces and angles. In 2000, Geoffrey Hinton described an imaging system [1] that combined the segmentation and recognition phase into a single inference process using parse trees. The credibility networks described the joint distribution over the state variables and over the possible parse trees. That system proved functional on the MNIST handwritten digit database. Capsule networks was introduced by Hinton and his teammates in 2017. The approach was claimed to reduce error rates on MNIST and to reduce training set sizes. Results were asserted to be considerably better than a CNN on highly overlapped digits. In Hinton's original idea one mini-column would entitled and detect one multidimensional entity. The most important advantage of using a summary is, it reduces the reading time.This output summary will be then be provided to the user in the user requested format I.e. either a voice output or a text file.

A Capsule Neural Network (CapsNet) is a machine learning system that is a type of artificial neural network (ANN) that can be used to better model hierarchical relationships. This approach is an attempt to closely mimic biological neural organization. Capsule is a nested set of neural layers. In a regular neural network you keep on adding more layers to it. In Capsule Network you would join more layers inside a single layer. Or In other words, nest a neural layer inside one another. The states of the neurons inside a capsule captures the above properties of one entity inside an image. A capsule thereby outputs a vector to represent the existence of a particular entity. A capsule is a group of neurons whose activity vector represents the instantiation measureable factors or parameters of a specific type of entity such as an object or an object part. We use the length of the activity vector to entitled the probability that the entity exists and its orientation to represent the instantiation parameters. Active capsules at one level make predictions or divination, via transformation matrices, for the instantiation parameters of higher-level capsules. When multiple predictions agree, a higher level capsule becomes active instantly. We show that a discriminatively trained and multi-layer capsule system achieves state-of-the-art performance on MNIST [2] and is considerably better than a convolutional net at identifying highly overlapping digits. To achieve these solutions we use a frequent iterative routing-by-agreement mechanism: “A lower-level capsule prefers to send its output to higher level capsules whose activity vectors have a huge scalar product with the prediction expected from the lower-level capsule.” Capsule networks try to imitate the way observations are made by humans onto computers. The authors state that human vision refuses irrelevant details by using a carefully determined sequence of fetish points, confirming that only a tiny fraction of the optic arrangement is ever processed at the highest resolution. The incentive stems from the fact that as of now, convolutional neural networks have problems modelling spatial relationships of parts making up an image. Instead of modelling co-existence and thereby disregarding relative positioning, capsule nets try to model the global relative transformations of different sub-parts along a hierarchy. Objective of the proposed system is to detect the object given as input, even if it is inverted or contains any impurity or noise within. This paper describes about the drawbacks of convolutional neural networks and of implementing Capsule network for overcoming the problems of Convolutional Neural Network.

II. Literature Review

In 1981, Geoffrey Hinton stated in his paper that there are many described that the first phase of the process involved training the neural networks and learn the types of images that should be included while detection.[3] The neural network learns the patterns from various images that should be included while detection and those that should not be included. This paper describes a way of using parallel hardware to implement a cooperative computation in which the process of choosing an object-based frame and the generation of a description relative to that frame occur simultaneously, with each influencing the other. The first phase of the process involves training the neural networks to learn the types of images that should be included while detecting. The neural network learns the patterns inherent in images that should be included while detection and those that should not be included. The neural network consists of seven input-layer of neurons, six hidden-layer of neurons, and one output layer of neuron. This is accomplished by the feature fusion phase, that consists of two steps: 1) eliminating uncommon features; and 2) disintegrating the effects of common features.

In 2011, A.Krizhevsky & S.D. Wang along with Geoffrey Hinton developed an auto-encoder [4] prediction which multiplies the probability of recognition of the capsule output. This Auto-encoder will tend to learn a representation in hidden layer that helps in rejection noise. In this paper a simple capsule-based network is used to imitate different viewing conditions of an implicitly defined visual entity. Each capsule results both the probability that a particular optical entity is present and a set of instantiation parameters like pose, lighting and deformation of the optical entity relative to a canonical version of that entity.

In 2017, Nicholas Frost & Sara Sabour developed Dynamic Routing between the capsules [5] such that the routers would be able to select the path according to real time logical Network layout changes. It would also be responsible for its creation, maintenance and updating as well. One of the advantages of capsules that output direct instantiation parameters is that they provide a simple way to recognize wholes by recognizing their parts. If a capsule can learn to show the pose of its visual or optical entity in a vector that is linearly related to the "natural" representations of pose used in computer graphics, there is a very simple and highly selective test for whether the visual entities represented by two active capsules, A and B, have the right spatial relationship to activate a higher-level capsule.

In 2018, Geoffrey Hinton, Nicholas Frost & Sara Sabour worked on Matrix Capsules by using EM Routing. This paper, titled "Matrix Capsules for Dynamic Routing," is widely speculated to have been authored by Hinton, and discusses a revolutionary new method for dynamic routing—even compared to his first paper. Essentially, when routing, we are making the assumption that each capsule in layer (l) is activated because it is part of some 'whole' in the next layer. In this step, we assume there is some latent variable that explains which 'whole' our information came from, and are trying to infer the probability that each matrix output came from the higher-order feature in level (l+1). This boils down to an unsupervised learning problem, and we tackle it with an algorithm called Expectation Maximization. In essence, expectation maximization attempts to maximize the likelihood that our data (outputs of layer (l)) is explained by the capsules in layer (l+1). Expectation Maximization is an iterative algorithm that consists of two steps: expectation, and maximization. Typically, for capsule networks, each step is performed three times, and then terminated.

III. Research Methodology

Our goal is to overcome the disadvantages of the convolutional neural network. The proposed system summarizes the given input which can be an digit between 0 - 9 which goes under process of prediction and classification. The given input image is then compared with the trained model sets and thus generating the output accordingly. If the given input is correct it will generate the output as authenticate data and if not then, it will prompt as null. The more advanced version of the system can be used in face detection and face-recognition as well.

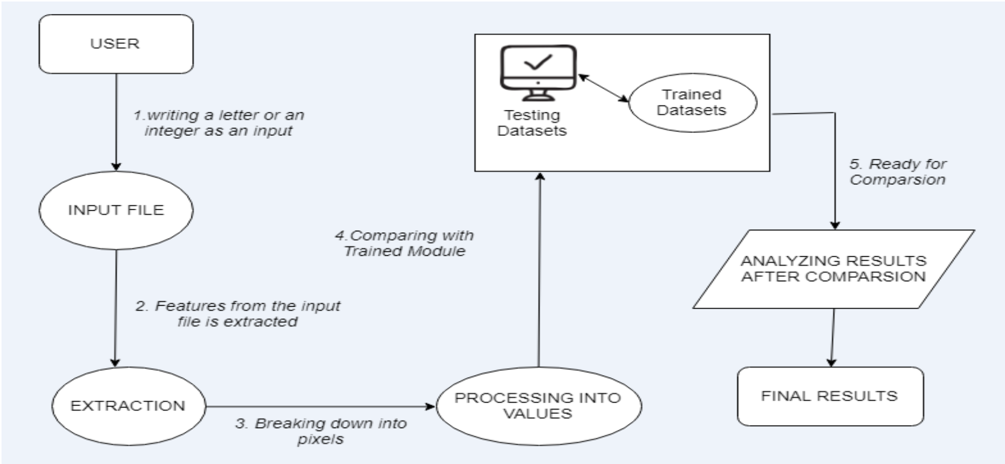

Fig 1. System Architecture

1. Input File

An input file will be an image of an animistic handwriting provided to the system. The file may either be in .jpg or .jpeg format. This file can be given in any position; it can be inverted as well tilted in certain angle.

2. Features Extraction

In machine learning, pattern recognition and analysis in image processing, feature extraction starts from an initial set of measured data and builds derived values -features intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps. It will reduce the amount of duplicate data for a given analysis by combining the variables into features. This helps in the process of detecting the image better.

3. Processing into values

For a grayscale images, the pixel value is a single number that represents the brightness of the pixel. The most common pixel format is the byte image, where this number is stored as an 8-bit integer giving a range of possible values from 0 to 255. Typically zero is taken to be black, and 255 is taken to be white. An image is more of data that numbers indicating variations of RGB values at a particular location on a grid of pixels.

4. Trained Datasets

Compare two datasets and summarise species occurrence and abundance of species recorded in dataset one across dataset two. Useful for examining the conformity between sediment core and training set species data. Overall using pre-trained models is effective at differentiating between the different types of objects, despite the fact that it hasn’t been trained any kind of images.

5. Authentication

Authentication is the act of confirming the truth of an attribute of a single piece of data (a datum) claimed true by an entity. In contrast with identification which refers to the act of stating or otherwise indicating a claim purportedly attesting to a person or thing's identity, authentication is the process of actually confirming that identity.

This is the data typically used to provide an unbiased evaluation of the final that are completed and fit on the training dataset. Actually, such data is used for testing the model whether it is responding or working appropriately or not.

6. Results

The output thus obtained is then displayed to the user. Detection Result will provide output whether the input image is detected or not. The output is thus displayed as requested by the user.

- Comparison of output with capsule network and convolutional network.
- Accuracy percentage will be shown for both the network.
- Overcome the drawbacks of convolutional neural networks.

IV. Experimental Results

The system implementation is divided into three phases:

1. Input given by the user which can be any integer or alphabet.
2. Prediction of the input using Convolutional network.
3. Prediction of the input using Capsule Network.

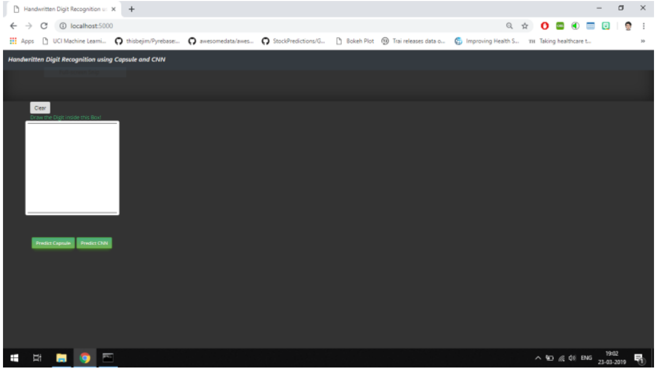

Fig 2: Home Page

The above figure is the home page of the system, from where we can draw any integer or alphabet on the blank notepad as an input to the system. Predict CNN will give you the output of prediction done by Convolutional Neural network. Predict Capsule will give the output of the prediction done by capsule network. Clear is for rewriting the input if user needs to change its input.

A. 4.1 System Flow

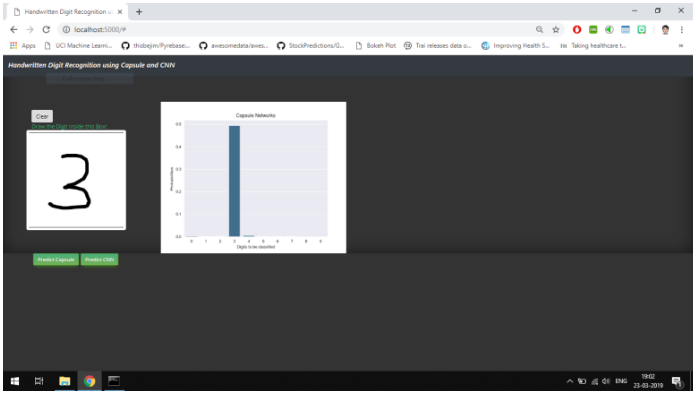

Fig 3 : Prediction done by Capsule Network

The above figure is the prediction done by Capsule network where the user has written 3 as an input to the system. After writing the input, user clicks on Predict Capsule and the system thereby will generate the output for it.

4.2 Comparison of output using capsule network and convolutional neural network

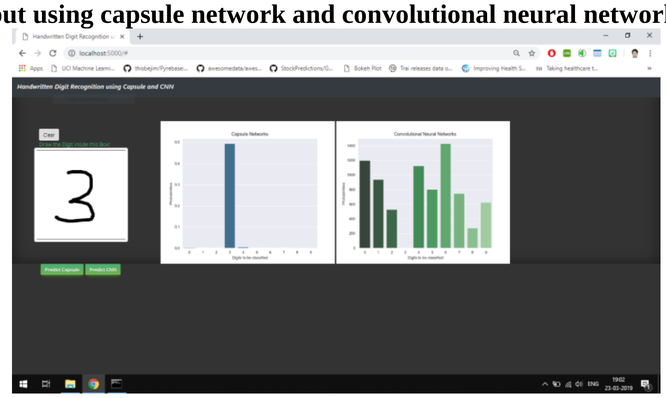

Fig 4 : Comparison with Capsule network and Convolutional Neural Network

The above figure shows the comparison of the output using both the networks. In the above figure we can see the accuracy of the input given to the system. Capsule network gives the accurate output while Convolutional neural network gives mixture of output. The reason why we see absolutely zero value in CNN graph is because we've trained the model with Negative Log Likelihood (NLL) Loss optimizer that computes with the difference of lowest value in graph with highest probability.

4.3 Simple input given to the system

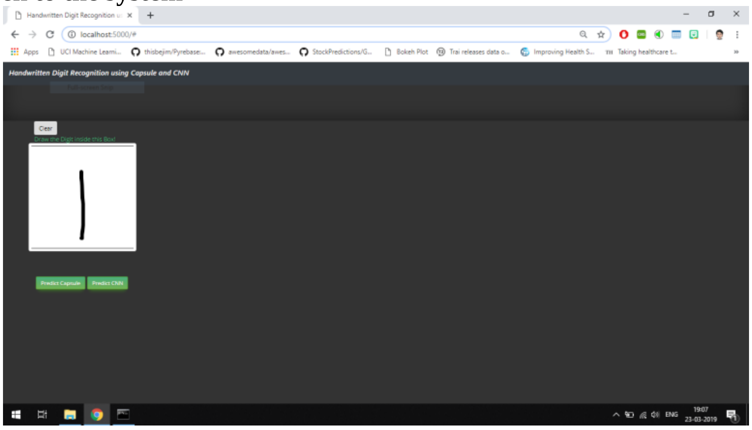

Fig 5 : Simple input sample

The above figure shows the input given to the system, which is an integer. i.e. 1. Now the user can see the comparison for both the network, for a human brain it is very easy to understand that the given input is 1. But the convolutional neural network still can't give the accurate answer.

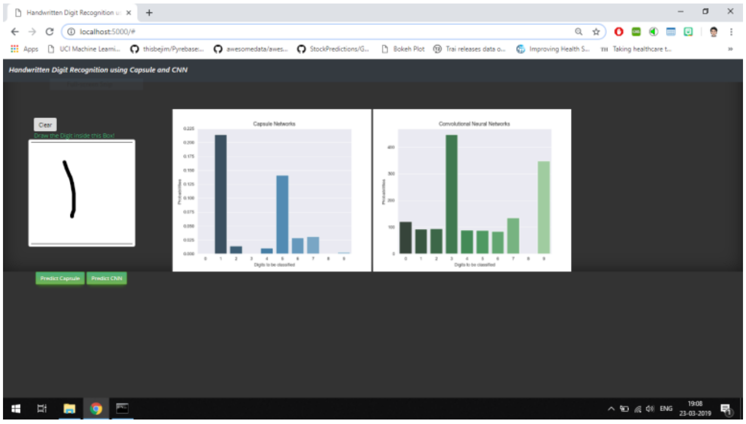

Fig 6 : Comparison for Integer one

The above figure shows the comparison of both the networks. Here the user can see that the result obtained by the capsule network gives better accuracy than the convolutional neural network.

V. Conclusion and Future Scope

This approach is to resolve the disadvantages of the convolutional neural network. Pooling problem is solved using EM Routing between the capsules [6]. Deriving the presence of object relationship is done by activation probability [7]. The noise problem [8] and the spatial understanding [9] is resolved by using Pose Matrix. So in comparison with the neural network, even if capsule network is complex still can resolve certain problems of neural network. Proposed system can be used where neural network fails.

References

[1]. Hinton, G.E., Ghahramani, Z and Teh Y. W.," Learning to parse images," Advances in Neural Information Processing Systems 12, MIT Press, Cambridge, MA:

[2]. Kussul, Ernst; Tatiana Baidyk, "Improved method of handwritten digit recognition tested on MNIST database," Retrieved 20 September 2013.

[3]. Hinton, G.E. ,"A parallel computation that assigns canonical object-based frames of reference," Proceedings of the Seventh International Joint Conference on Artificial Intelligence Vol 2, Vancouver BC, Canada

[4]. Hinton, G. E., Krizhevsky, A. and Wang, S ,"Transforming Auto-encoders.," International Conference on Artificial Neural Networks, Helsinki :http://www.cs.toronto.edu/~hinton/absps/transauto6.pdf

[5]. Sabour, S., Frosst, N. and Hinton, G. E, "Dynamic Routing between Capsules," NIPS-2017: http://www.cs.toronto.edu/~hinton/absps/DynamicRouting.pdf

[6]. Jay Ricco, Fourth Year, Undergraduate BS Computer Science:https://www.quora.com/What-is-max-pooling-in-convolutional-neural-networks Accessed on 15.11.2018,8 PM

[7]. Alexey Tarasov Felix Kling ,"knowledge discovery and data mining," : https://www.cs.uic.edu/~urbcomp2013/papers/Paper 18_Alexey Tarasov.pdf

[8]. Robert H. Kingston," Detection of Optical and Infrared Radiation,": https://link.springer.com/chapter/10.1007/978-3-540-35948-7_2

[9]. "Understanding of Coordinate Reference Systems,": https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/coordinate_reference_systems.html

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Shraddha More. " Overcoming the Drawbacks of Convolutional Neural Network Using Capsule Network" IOSR Journal of Computer Engineering (IOSR-JCE) 21.2 (2019): 06-11.